知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于自述研究兴趣相似性网络的机构潜在合作关系挖掘

——以国内图书情报与档案管理教育机构为例

胡志伟 裴雷
南京大学信息管理学院 南京 210023

摘要：[目的 / 意义] 定量描述图书情报与档案管理学科的研究图景，为各机构之间合作关系的建立提供决策支持，从而推动跨机构合作的发展。[方法 / 过程] 采用 LDA 主题模型和网络分析方法，以国内 67 所图书情报与档案管理教育机构为例，通过对教师自述研究兴趣文本进行主题聚类构建机构相似性网络，并进行社群划分与潜在合作关系挖掘。[结果 / 结论] 当前国内图书情报与档案管理教师的研究兴趣主要涉及信息资源管理、信息计量与竞争情报、信息服务与用户等 11 个主题，样本机构可划分为 7 个社群，包含 457 对潜在合作关系。未来，图书情报与档案管理学科除了向 5 种路径进行学科融合之外，还可在不同领域充分展开科研与教育实践的跨机构合作。

关键词：相似性网络　自述研究兴趣　科研合作　教育机构　LDA 模型

分类号：G203

引用格式：胡志伟 , 裴雷 . 基于自述研究兴趣相似性网络的机构潜在合作关系挖掘 : 以国内图书情报与档案管理教育机构为例 [J/OL]. 知识管理论坛 , 2022, 7(2): 143-152[引用日期]. http://www.kmf.ac.cn/p/282/.

1 引言

2019 年 4 月，为提高高校服务经济社会发展能力，教育部、中央政法委、科技部等 13 个部门联合启动“六卓越一拔尖”计划 2.0，该计划包括全面推进新文科建设 [1]。较传统文科而言，新文科更加强调学科发展中的继承与创新、多学科的交叉与融合以及多机构的协同与共享 [2]。在大数据、云计算、人工智能等新技术高速发展和信息社会需求不断变化的时代背景下，图书情报与档案管理学科展现了日益显著的跨学科特性，而学科的多向发展也对跨系统组织之间的科研合作提出了更多的要求。跨机构合作能够在信息共享的基础上充分利用科研资源，形成解决复杂研究问题的有效方法 [3]。如何挖掘机构之间的潜在合作关系、提升跨机构合作效率，也成为了科学研究中的重要议题。

作者简介：胡志伟，硕士研究生，E-mail: mg1914013@smail.nju.edu.cn；裴雷，教授，博士生导师。

收稿日期：2021-10-27　　发表日期：2022-03-24　　本文责任编辑：刘远颖

一般认为，研究内容更为相近的实体之间具有更高的潜在合作程度。对此，相关研究[4-6]主要从学者的科研成果中直接获取或间接识别研究主题，并采用网络分析等方法，从不同角度对科研实体进行相似度测算或社群分析。然而，此条研究路径很少利用公开于互联网中的官方自述资料。当前以自述资料为数据来源的研究[7-9]主要是对主题分布情况的内容分析，而基于主题模型的方法存在一定程度的缺位。鉴于此，笔者拟以国内图书情报与档案管理教育机构为例，基于学者的自述研究兴趣文本，利用隐含狄利克雷分布（Latent Dirichlet Allocation，LDA）主题模型和网络分析方法，试图回答如下两个问题：

（1）当前国内图书情报与档案管理专业教师的研究兴趣集中表现为哪些主题?

（2）国内的哪些图书情报与档案管理教育机构之间存在潜在的合作基础?

本文的研究结果有助于理解国内图书情报与档案管理专业教师研究兴趣的分布情况，协助各机构在做出战略定位和规划的基础上寻求同自身教学科研实践内容相似的对应机构，为其间合作关系的建立提供决策支持。

2 相关研究

2.1 学者研究兴趣主题挖掘

学者的研究兴趣代表了学者在一定时间段内感兴趣的研究内容，可通过学术成果（包括论文、项目、专利等）[10]、科研社交网站[11]、个人主页[12]等途径获取相关数据。由于学者与研究兴趣之间往往是多对多的关系，即一名学者会对多个研究主题感兴趣，且不同的学者会同时对某项研究主题感兴趣，因此许多研究聚焦于学者研究兴趣的主题识别与表征工作。当前研究主要采用主题模型和知识网络方法实现上述目标，例如刘晓豫等[13]利用加权 K-means 聚类算法对多专长专家研究兴趣的识别方法的改进，熊回香等[14]利用 LDA 主题模型对学者主题档案的建构，刘萍等[15]利用关键词共现网络对特定机构学者的社区划分等。值得注意的是，当前相关研究主要将科研成果的题录信息作为底层数据，而很少针对学者自述资料进行分析。一般而言，学者自述研究兴趣在表达凝练的同时蕴含着丰富语义内容，且在时效性得到保证的前提下，能够有效减小科研成果发表周期较长、与学者实际研究兴趣关联程度较低所带来的影响。因此，有必要对学者的自述资料给予一定的重视。

2.2 科研实体潜在合作关系挖掘

为了推动潜在科研合作以提高科研效率，研究者对科研实体之间的相似性进行了探讨。X. Kong 等[16]基于 LDA 主题模型所揭示的学者动态研究兴趣矩阵，通过计算学者个人之间的余弦相似度构建了合作者推荐模型；安璐等[17]基于“机构—文献类别”矩阵，利用自组织映射方法测算了中美图书情报科研机构研究领域的相似性；林原等[18]基于“高校—合作”和“高校—主题”矩阵，分别借助改进后的 Katz 和余弦相似性指标，对高校之间的潜在合作关系进行了分析；E. Yan 等[19]基于 SimRank 模型，对学者、机构和国家 3 个层次的合作网络进行了链路预测。总体而言，在机构相似性测度层面，相关研究主要利用机构主题模型和已有合作网络实现测算，其中余弦相似度在关于前者的研究中得到了广泛应用。笔者以国内图书情报与档案管理教育机构为例，借助 LDA 主题模型对专职教师的自述研究兴趣进行主题挖掘，从而依次构建机构的主题向量模型和余弦相似性网络，探测并分析其间的潜在合作关系。

3 数据与方法

3.1 研究设计

笔者主要采用网络调查、LDA 主题聚类和网络分析方法，在 R 语言和 Gephi 软件环境中完成以下研究路径（见图 1）：①数据获取与预处理。通过样本机构的官方网站采集教师的个人资料，经文本数据清洗和分词操作后，构建“教

知识管理论坛，2022（2）:143-152
DOI: 10.13266/j.issn.2095-5472.2022.012

师—词项”DTM（Document Term Matrix）矩阵。②教师自述研究兴趣主题发现。确定 LDA 主题模型的各项基本参数，根据“主题—词项”分布情况对主题发现结果进行标识。③机构相似性网络分析。基于机构的主题向量模型计算机构之间的相似度，从而构建机构的相似性网络，对机构进行社群划分，识别具有潜在合作关系的机构组。

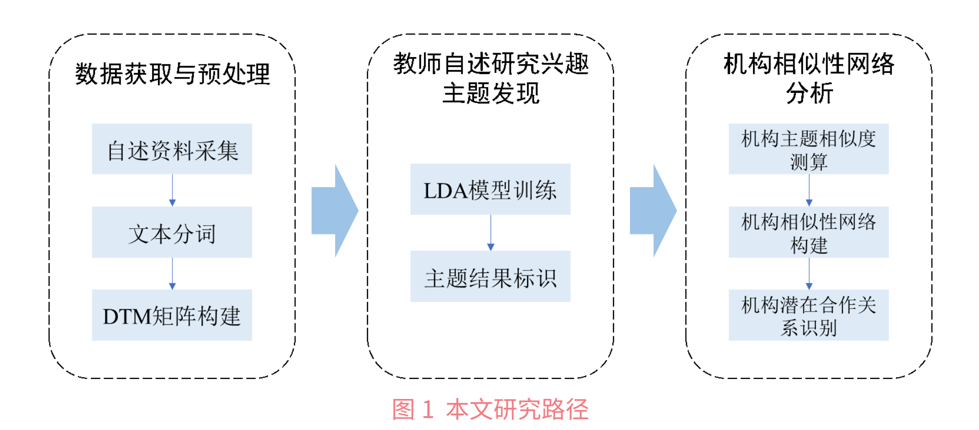


图 1 本文研究路径

3.2 数据获取与预处理

经过对国内图书情报与档案管理教育机构官方网站建设现状的调查，笔者最终选取了 67 所机构作为研究的数据来源（为行文方便，下文仅以机构所属高校或科研单位名称表述机构全称）。在此基础上，以官方网站为入口采集教师的个人资料，涉及姓名、所属机构、自述研究兴趣等方面。数据采集时间为 2020 年 11 月 16 日至 11 月 29 日，经清洗后共获得 1 289 名教师的个人信息，提及研究兴趣 3 599 次。

分词处理与矩阵构建操作主要借助 R 语言中的 Rwordseg 包和 tm 包完成。首先，因样本中存在中英文多语种来源，笔者对英文描述进行了英译中处理；其次，为了删除无意义的词语，采用哈尔滨工业大学停用词表和四川大学机器智能实验室停用词库构建分词停用词表；最后，抽取 20% 的样本进行预分词实验，结合实际情况补充了自定义词典，其主要涉及学科名称等专有名词。经过上述操作后，得到了 1289 × 14179 的“教师—词项”DTM 矩阵。

3.3 关键方法

3.3.1 LDA 主题模型

为分析国内图书情报与档案管理专业教师的研究兴趣主题分布，需要确定 LDA 主题模型的基本参数。LDA 主题模型的构建主要借助 R 语言中的 lda 包、Rmpfr 包、topicmodels 包和 LDAvis 包实现。笔者主要采用 Gibbs 采样算法求解全局主题和词项的分布，需设定 α、β 和最优主题数 K 3 个超参数。其中，α、β 一般设定为默认值，可根据实际效果进行调节。可根据困惑度或似然估计数值随主题数变化的情况确定，当似然估计数值达到最大时的主题数即为最优解。在此基础上，从模型训练所得的“主题—词项”概率分布矩阵中抽取各主题下具有较高显著度 [20] 的若干词项，作为主题标识的依据。此外，基于“教师—主题”概率分布矩阵计算所有教师的研究兴趣涉及某一主题的概率和占总体的比重，可以得到该主题的强度，作为主题关注度的计量指标。

3.3.2 相似度计算与分析

笔者基于机构的主题向量模型，选取余弦相似度作为机构间相似性测度的指标。首先，根据 LDA 主题模型训练所得的“教师—主题”概率分布矩阵，可得到各教师的主题向量模型 V_{F_i}，计算公式如公式（1）所示：

$$V_{F_i}=(P_{i,1},P_{i,2},\cdots,P_{i,K}), i=1,2,\cdots,1289, j=1,2,\cdots,K \quad \text{公式（1）}$$

其中，P_{ij} 为教师 F_i 的研究兴趣文本隶属于

主题 T_j 的概率，K 为最优主题数。分别计算所属机构为 I_x 的所有教师 $F_{x,i}$ 的主题概率分布和，从而构建各机构的主题向量模型 V_{I_x}，计算公式如公式（2）所示：

$$V_{I_x}=\sum_{i=1}^{n}V_{F_{x,i}},x=1,2,\cdots,67,i=1,2,\cdots,n \quad 公式(2)$$

其中，n 为机构 I_x 的教师总数。因此，机构 I_a 和 I_b 之间的余弦相似度计算公式可以表示为：

$$Sim(I_a,I_b)=\frac{V_{I_a}\times V_{I_b}}{\sqrt{V_{I_a}^2}\times\sqrt{V_{I_b}^2}},a,b=1,2,\cdots,67,$$
$$0<Sim(I_a,I_b)\le 1 \quad 公式(3)$$

显然，$Sim(I_a,I_b)=Sim(I_b,I_a)$，且当 $a=b$ 时，$Sim(I_a, I_b)=1$。一方面，可据此构建机构之间的相似性矩阵。由于基于此矩阵生成的网络为连通图，为便于后续分析，需要根据一定策略设定合适的阈值 λ 进行剪枝操作，剔除权重较低的矩阵元素。笔者参照帕累托法则，仅保留权重较高的 20% 的元素，借助 Gephi 软件构建机构之间的相似性网络，并使用内置的 Louvain 算法 [21] 作为社群划分方法。此网络的节点为各机构，连接节点 I_a 和 I_b 的边的权重为 $Sim(I_a, I_b)$，代表了各机构之间的相似程度。另一方面，可分别得到某机构和与其最为相似的机构主题强度超过 10% 的主题集合，通过计算交集获取两者的潜在合作研究领域。

4 研究结果

4.1 教师自述研究兴趣主题发现

首先，参照 LDA 主题模型基本参数的确定方法，确定聚类的最优主题数。如图 2 所示，当主题数为 11 时达到最大似然估计，说明此时模型性能较好，能够有效地实现主题的判别，因此设定最优主题数 K=11。在运行主题发现模型得到“主题—词项”概率分布结果后，依据每个主题下所包含的显著度最高的 6 个词项对主题进行标识（见表 1）。目前，国内图书情报与档案管理专业教师的研究兴趣可主要概括为电子商务与战略管理、信息服务与用户、图书馆学与图书文化史等 11 个主要主题。

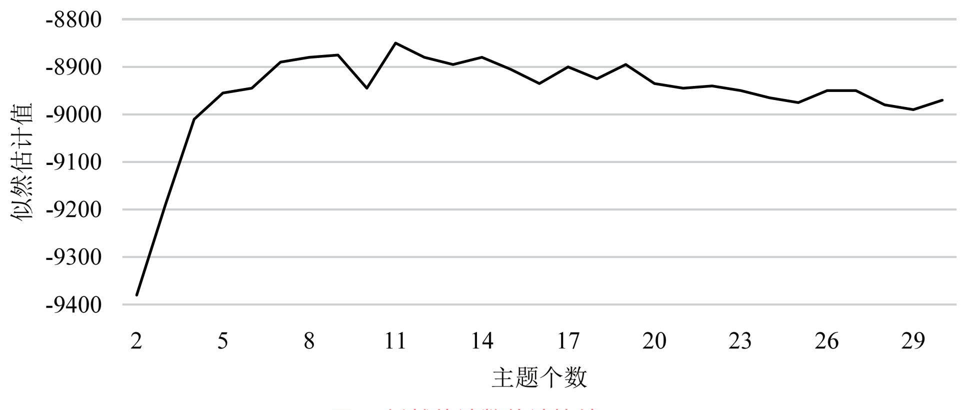


图 2 似然估计数值计算结果

如前所述，可基于 LDA 主题发现所得的“教师—主题”概率分布矩阵，计算某一主题占总体的比重，得到该主题的强度。在一定程度上，主题强度能够体现当前图书情报与档案管理专业教师对不同研究兴趣的关注情况，见图 3。信息资源管理（T_8）、信息计量与竞争情报（T_7）以及信息服务与用户（T_2）是样本教师群体最为关注的 3 个主题，其主题强度均超过 10%。这说明在当前时间窗口下，国内图书情报与档案管理专业教师对学科的内核形成了初步共识，并集中体现了学科整体面向国家战略、服务社会发展、满足用户需求的使命感。国内图书情报与档案管理专业教师的研究兴趣也体现了较高的技术介入程度与较广的学科应用场景，电子商务与战略管理

知识管理论坛，2022（2）:143-152
DOI: 10.13266/j.issn.2095-5472.2022.012

（T_1）、数据挖掘与大数据技术（T_{11}）和政务信息管理（T_9）呈现了相对较高的主题强度。此外，出版与文化产业（T_5）、信息跨学科应用（T_4）以及图书馆学与图书文化史（T_3）是关注度相对较低的 3 个主题，这主要受样本机构相应学科的建设规模与师资数量所影响。

表 1 教师自述研究兴趣的“主题—词项”分布

主题序号	主题标识	高显著度词项
T_1	电子商务与战略管理	电子 管理 商务 系统 企业 战略
T_2	信息服务与用户	信息 分析 网络 用户 服务 行为
T_3	图书馆学与图书文化史	历史 图书馆 文献 文化 阅读 古籍
T_4	信息跨学科应用	医学 资源 信息学 生物 健康 组织
T_5	出版与文化产业	出版 文化 公共 产业 阅读 营销
T_6	管理科学与工程	管理 系统 供应链 物流 经济 工程
T_7	信息计量与竞争情报	情报 信息 科技 评价 竞争 计量
T_8	信息资源管理	信息 管理 资源 组织 检索 建设
T_9	政务信息管理	管理 档案 电子 政务 政府 治理
T_{10}	档案管理理论与方法	档案 档案学 理论 管理 利用 文件
T_{11}	数据挖掘与大数据技术	数据 挖掘 智能 网络 大数据 技术

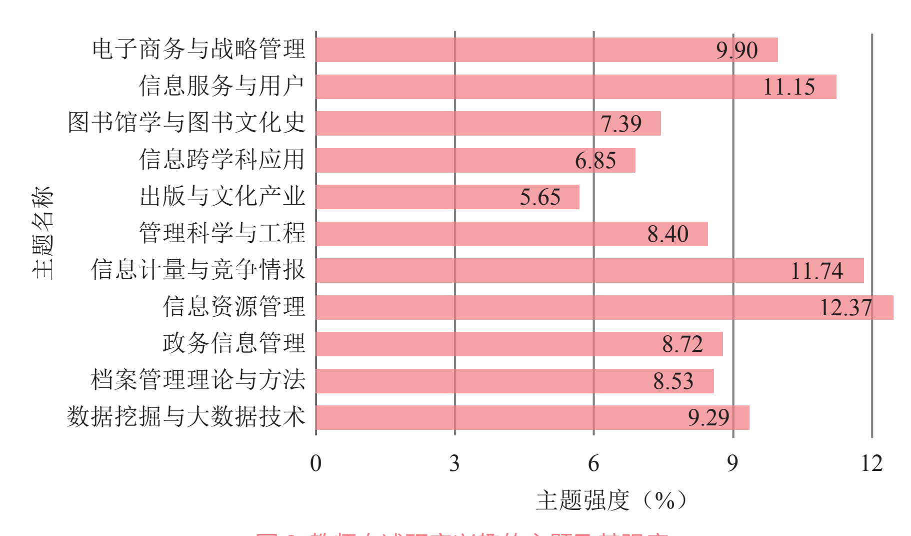


图 3 教师自述研究兴趣的主题及其强度

4.2 机构相似性网络分析

在获取机构的主题向量模型后，可以计算出机构之间的余弦相似度。根据帕累托法则对机构之间的边进行剪枝操作，此时权重阈值 λ=0.72，即当 $Sim(I_a, I_b)\geq 0.72$ 时，可以认为机构 I_a 和 I_b 之间具有高度的相似性。在此基础上，可以得到国内图书情报与档案管理机构的相似性网络（见图 4），共包含 67 个节点、457 条边。其中，节点大小代表机构师资数量的多少，节点颜色代表不同的社群类别，分别以字母 A-G 表示。

社群 A 主要由以档案学或信息资源管理专业为主体的机构组成，包括中国人民大学、上

海大学、湘潭大学、郑州大学等 23 所机构。档案管理理论与方法、政务信息管理和信息资源管理是社群 A 较为关注的 3 个主题，三者的强度和达到了 56.6%。

图 4 国内图书情报与档案管理教育机构相似性网络

社群 B 内机构的学科背景较为综合全面，与社群 A 构成了网络的主体部分，包括武汉大学、南京大学、北京大学、中山大学等 19 所机构。社群 B 内部的主题概率分布相对均衡，信息资源管理、信息服务与用户、信息计量与竞争情报以及数据挖掘与大数据技术是其较为关注的 4 个主题，强度均超过了 10%。

社群 C 主要由具有经济管理背景的机构组成，多设于经济或管理学院（部）下，包括河海大学、山东大学、山西大学、东南大学等 10 所机构。强度较高的两个主题为电子商务与战略管理以及管理科学与工程，均在 20% 以上。

社群 D 包括中国科学技术信息研究所、中国科学院大学、江苏大学、山东理工大学等 9 所机构，主要依托机构图书馆或文献情报中心进行学科建设。信息计量与竞争情报在该社群内的强度最高，达到了 43.4%。具备图书资料专业职称的教职工约占社群内教师总体的 64.3%，多在科研评价、科技战略和专利情报分析等方向展开研究。

社群 E、F 和 G 的机构数相对较少，且与其他社群之间几乎不具有显著的相似性。社群

E 包括中国医科大学、山西医科大学和中南大学 3 所机构，均具有医学或生物学背景，信息跨学科应用的主题强度高达 65.4%。社群 F 包括复旦大学、中国中医科学院 2 所机构，尽管同社群 D 一样依托图书资料中心建设，但其对历史文献学和古籍保护方向给予了更多的强调，图书馆学与图书文化史的主题强度为 38.2%，显著高于社群 D 的 1.1%。社群 G 仅包括山东科技大学 1 所机构，为孤立节点，同其他 66 所机构均不连通。这主要缘于其深厚的计算机技术背景，其数据挖掘与大数据技术的主题强度高达 68.0%，而其他机构这一主题的最高强度也不足 30%。

在前文识别457对潜在合作关系的基础上，寻找与某一机构最为相似的其他机构，并识别两者合作的潜在研究领域，对推动跨机构合作实践有着重要的意义。笔者将某机构和与其最为相似的机构主题强度超过 10% 的主题集合定义为潜在研究领域，表 2 展示了部分计算结果。在相似度数值方面，东南大学与河海大学之间的相似程度最高，达到了 0.98，两者对数据挖掘与大数据技术以及管理科学与工程方向给予了高度关注；中国中医科学院的最相似机构为复旦大学，相似度为 0.76，两者的潜在合作领域分别为信息计量与竞争情报以及图书馆学与图书文化史。此外，表 2 展示了学科建设水平位列前位的机构相关情况，如南京大学可在信息服务与用户、数据挖掘与大数据技术以及出版与文化产业等领域与武汉大学开展合作，而中国人民大学可在信息资源管理、政务信息管理以及档案管理理论与方法等领域与北京联合大学开展合作等。

表 2 同各机构相似度最高的其他机构与潜在合作领域（部分结果）

相似度排序	机构名称	最相似机构名称	相似度	潜在合作领域
1	东南大学	河海大学	0.98	数据挖掘与大数据技术、管理科学与工程
20	武汉大学	中山大学	0.93	信息资源管理、信息服务与用户、电子商务与战略管理、数据挖掘与大数据技术
36	南京大学	武汉大学	0.91	信息服务与用户、数据挖掘与大数据技术、出版与文化产业
49	中国人民大学	北京联合大学	0.89	信息资源管理、政务信息管理、档案管理理论与方法
54	北京大学	南京农业大学	0.88	信息资源管理、信息服务与用户、信息跨学科应用
60	华中师范大学	西安电子科技大学	0.84	信息服务与用户、数据挖掘与大数据技术、管理科学与工程
66	中国中医科学院	复旦大学	0.76	信息计量与竞争情报、图书馆学与图书文化史

5 讨论

5.1 图书情报与档案管理的跨学科融合路径

作为一门应用型学科，图书情报与档案管理的学科发展具有明显的跨学科特征，一般认为学者的跨学科研究活动是其跨学科的重要表现和驱动。基于前文对我国专职教师研究兴趣的主题发现和教育机构的社群划分，笔者总结了图书情报与档案管理的跨学科融合路径，见图 5。

一方面，图书情报与档案管理内部应紧密围绕信息资源管理这一学科内核发展。每个学科都有属于自身的话语系统和规则，在与其他学科融合的过程中，仍需明确并坚守自身的核心内容。图书情报与档案管理的核心知识即立足于信息资源全生命周期管理的教育与研究，涉及数据、信息和知识的记录、保存、组织、检索、获取、分析和开发利用过程，具有显著的服务特征 [22]。通过前文的主题发现，信息资

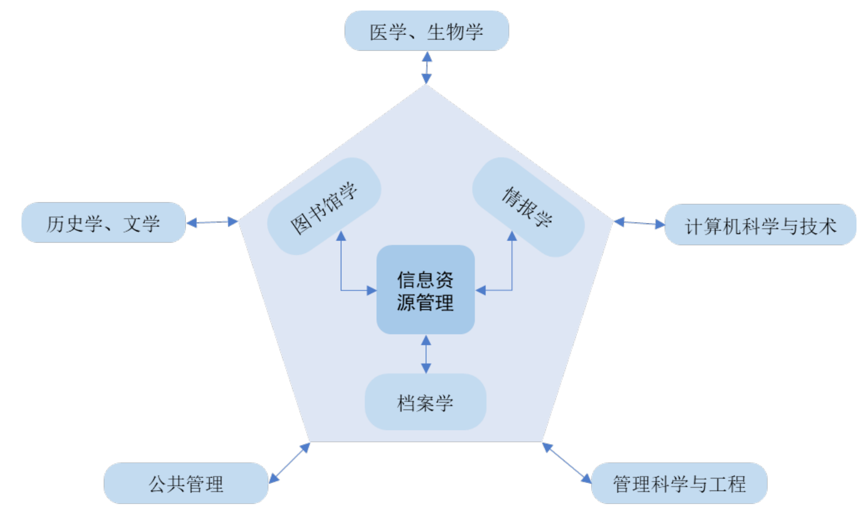


图 5 图书情报与档案管理的跨学科融合路径

源管理在当前国内专业教师的研究兴趣中具有最高的主题强度，且随着信息社会的不断发展，在大数据、人工智能等新兴信息技术的支持作用下，应用场景不断得到拓展与深化。在 2021 年的新版学科专业目录草案中，“图书情报与档案管理”一级学科拟更名为“信息资源管理”，进一步引发了有关学科核心知识的讨论。而这一行动充分体现了我国图书情报与档案管理一级学科的价值转向。此外，信息计量与竞争情报、信息服务与用户、电子商务与战略管理以及数据挖掘与大数据技术等将“数据—信息—知识—智慧”（DIKW）体系各层次之间联系起来的主题的强度亦相对较高，充分反映了图书情报与档案管理学科的整体联系更加紧密。图书馆学、情报学和档案学各二级学科的科研或教育实践者应树立学科共同体意识，在专注于信息资源管理内核的基础上，强化学科的社会服务职能，推动社会的创新与进步。

另一方面，图书情报与档案管理可通过同其他学科的交叉和融合促进学科深度发展，提升自身的学科地位与竞争力。随着当代科学的学科整合趋势愈加显著，图书情报与档案管理由于本就具有相对较强的跨学科属性，因而当其在对二级学科进行整合、凝聚一级学科内核的基础上，也在与同族学科之外的学科开展交融。通过上一节的研究主题发现与社群分析，可以发现国内图书情报与档案管理目前至少与 5 类信息资源管理学科群之外的学科实现了不同程度的融合，具体而言：与历史学、文学等基础人文学科相融合，助力图书文化史、数字人文等领域的研究；与医学、生物学等基础自然科学融合，用图书情报学的方法梳理文献和知识；与计算机科学与技术融合，借助新兴信息技术提升图书情报科研工作的效率；与管理科学与工程融合，为社会各行业或领域提供管理决策支持；与公共管理学科融合，推动政务信息管理、公共文化方向的发展。未来，图书情报与档案管理仍可继续深化与其他学科的融合与合作，更加开放地加强学科建设，敢于在跨学科研究中发声、竞争与引领。

5.2 新文科建设背景下的跨机构合作

机构之间的协同与信息共享是新文科建设所强调的一个重要方面。在多学科交叉融合的基础上，各机构能够通过跨机构的科研与教育合作有力推动自身转型，构建学科命运共同体。前文基于教师自述研究兴趣文本的余弦相似性网络将国内图书情报与档案管理教育机构划分为 7 个社群，并探测了与各机构最相似的潜在合作机构和研究领域。结果显示，大部分机构的潜在合作机构均处于相同社群中，且不同社

知识管理论坛，2022（2）:143-152
DOI: 10.13266/j.issn.2095-5472.2022.012

群的机构依托各种类型的教育与科研资源，形成了不同的研究侧重点以及鲜明的研究特色。例如，具有经济管理背景的机构在电子商务与战略管理以及管理科学与工程两个主题领域占据绝对优势，而依托于机构图书馆或文献情报中心开展学科建设的机构多专注于信息计量与竞争情报主题的研究。随着信息交流障碍的日益减少，国内图书情报与档案管理院系能够与相似机构之间就同一主题，在教育和科研等诸多场景中展开交流、分享与合作。

在此基础上，在本文基于研究主题分布的相似程度而得出的最显著潜在机构合作关系中，“复旦大学—陕西理工大学”“河北大学—武汉大学”“华东师范大学—南京大学”“吉首大学—长春师范大学”以及“南开大学—中山大学”为 5 组跨社群组合。这在一定程度上说明经识别所得的 7 个社群并非完全隔绝，存在相互渗透与相互借鉴的可能性。在这一过程中，处于社群边缘的机构扮演了重要的桥梁角色，它们往往具备着复合型的社群特征，能够有效地促进图书情报与档案管理学科内部的交流与合作，如河北大学、天津师范大学、山东大学、南开大学和四川大学等。未来，在发挥具备较大师资体量的机构引领学科发展作用的同时，也需注重发挥典型机构在科研与教育实践合作过程中的作用。此外，由于各社群间的机构数量不一，甚至在机构相似性网络中出现了山东科技大学这一依托于计算机科学与技术学科的孤立节点，因此在具体合作过程中，仍需注意研究内容与主题的分工，最大化图书情报与档案管理学科内部的凝聚力，共同促进合作效率的提升与研究成果的转化。

6 结语

笔者基于国内图书情报与档案管理教育机构的教师自述研究兴趣文本，对当前国内相关专业教师的研究兴趣进行主题聚类，并在此基础上构建机构的研究兴趣相似性网络，识别各机构的潜在合作机构及研究领域。就国内的图书情报与档案管理学科而言，研究发现：①教师的研究兴趣主要涉及信息资源管理、信息计量与竞争情报、信息服务与用户等 11 个主题；②样本院系主要可划分为 7 个社群，包含 457 对潜在合作关系。据此，笔者强调多学科融合与跨机构合作的重要性：①图书情报与档案管理的跨学科融合路径应在坚守信息资源管理内核的前提下，与管理科学与工程、公共管理、计算机科学与技术等学科积极融合；②充分发挥社群边缘机构在跨机构合作中的重要作用，促进机构之间多方面、深层次的合作。本文可以协助各机构寻求同自身相似的同行机构，并能够通过挖掘其间的潜在合作关系，在一定程度上推动跨机构合作的发展。

需要注意的是，本文的研究兴趣文本数据主要来自机构官方网站，可能存在数据老旧、更新不及时的情况，影响分析结果。今后，可综合机构网站、学术成果题录等多来源资料，优化数据质量，进行综合分析。此外，潜在合作关系与实际合作现状之间的比较、全球范围内机构潜在合作关系的探究，也是值得进一步研究的方向。

参考文献：

[1] 人民网 . “ 六卓越一拔尖 ” 计划 2.0 掀起一次中国高教 “ 质 量 革 命 ”[EB/OL]. [2021-01-30]. http://edu.people.com.cn/n1/2019/0430/c1006-31059188.html.

[2] 马费成 , 李志元 . 新文科背景下我国图书情报学科的发展前景 [J]. 中国图书馆学报 , 2020, 46(6): 4-15.

[3] LEE S, BOZEMAN B. The impact of research collaboration on scientific productivity[J]. Social studies of science, 2005, 35(5): 673-702.

[4] 刘萍 , 郭月培 , 郭怡婷 . 利用作者关键词网络探测作者相似性 [J]. 现代图书情报技术 , 2013(12): 62-69.

[5] 江文华 , 徐健 , 李纲 , 等 . 基于研究兴趣相似性网络的我国图书馆学研究社群分析 [J]. 现代情报 , 2019, 39(9): 21-27.

[6] 林原 , 王凯巧 , 刘海峰 , 等 . 网络表示学习在学者科研合作预测中的应用研究 [J]. 情报学报 , 2020, 39(4): 367-373.

[7] 史雅莉 , 司莉 , 姚瑞妃 , 等 . 国外 iSchools 图情学科教师队伍结构与研究领域调研及启示 [J]. 图书情报知识 , 2018(5): 33-40.

[8] 赵跃，梅梦娜．我国档案学研究的坚守、转向与进路——基于档案学者研究方向变化的审视 [J]. 档案学通讯，2019(3): 4-11.

[9] 肖希明，杨蕾．我国 LIS 教育机构师资队伍建设的调查分析 [J]. 图书情报知识，2014(1): 4-10.

[10] 陈翀，李楠，梁冰，等．基于成果特征的学者学术专长识别方法 [J]. 图书情报工作，2019, 63(20): 96-103.

[11] COPIELLO S. Research interest: another undisclosed (and redundant) algorithm by ResearchGate[J]. Scientometrics, 2019, 120(6): 351-360.

[12] HOLMBERG K, TSOU A, SUGIMOTO C R. The conceptual landscape of iSchools: examining current research interests of faculty members[J]. Information research, 2013, 18(3): 19-22.

[13] 刘晓豫，朱东华，汪雪锋，等．多专长专家识别方法研究——以大数据领域为例 [J]. 图书情报工作，2018, 62(3): 55-63.

[14] 熊回香，杨雪萍，蒋武轩，等．基于学术能力及合作关系网络的学者推荐研究 [J]. 情报科学，2019, 37(5): 71-78.

[15] 刘萍，周梦欢．基于共词网络的专家专长挖掘 [J]. 情报科学，2012, 30(12): 1815-1819.

[16] KONG X, JIANG H, WANG W, et al. Exploring dynamic research interest and academic influence for scientific collaborator recommendation[J]. Scientometrics, 2017, 113(1): 369-385.

[17] 安璐，余传明，李纲，等．中美图情科研机构研究领域比较研究 [J]. 中国图书馆学报，2014, 40(5): 64-77.

[18] 林原，谢张，李鲁萤，等．卓越大学联盟国内外高校科研合作机会发现 [J]. 现代情报，2020, 39(3): 81-86, 114.

[19] YAN E, GUNS R. Predicting and recommending collaborations: an author-, institution-, and country-level analysis[J]. Journal of informetrics, 2014, 8(2): 295-309.

[20] CHUANG J, MANNING C D, HEER J. Termite: visualization techniques for assessing textual topic models[C]//Proceedings of the International Working Conference on Advanced Visual Interfaces. New York: ACM, 2012: 74-77.

[21] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics: theory and experiment, 2008(10): P10008.

[22] 闫慧．青年学者论图情档一级学科核心知识及发展方向：2019 年图书情报与档案管理青年学者沙龙会议述评 [J]. 中国图书馆学报，2019, 45(1): 121-127.

作者贡献说明：

胡志伟：收集、整理并分析数据，撰写论文；

裴　雷：指导研究思路，核查论文内容并提出修改意见。

Mining Potential Cooperative Relationships Between Institutions Based on Similarity Network of Self-Reported Research Interests: A Case Study of Library, Information and Archives Management Schools in China

Hu Zhiwei Pei Lei

School of Information Management, Nanjing University, Nanjing 210023

Abstract: [Purpose/Significance] By quantitatively describing the research landscape of the discipline of library, information and archives management (LIAM), this paper can provide decision support for the establishment of cooperative relationships between institutions, thus promoting the development of inter-institutional cooperation. **[Method/Process]** By using LDA model and network analysis method, this paper took 67 LIAM schools in China as an example. By thematic clustering of faculty's self-reported research interest texts, the similarity network of institutions was constructed, and community division and potential cooperative relationship mining were performed. **[Result/Conclusion]** It is found that the current research interests of LIAM faculties in China mainly involve 11 topics, such as information resource management, informetrics and competitive intelligence, and information services and users. The sample schools can be divided into 7 communities, containing 457 pairs of potential cooperative relationships. In the future, in addition to taking part in five approaches of disciplinary integration, LIAM can fully carry out inter-institutional cooperation in scientific research and educational practice in different fields.

Keywords: similarity network　self-reported research interests　scientific cooperation　educational institution　LDA